ARI Research Note 87-80

DTIC FILE COPY

AD-A188 655

VISUAL DISPLAY REPRESENTATION OF MULTIDIMENSIONAL SYSTEMS: THE EFFECT OF INFORMATION CORRELATION AND DISPLAY INTEGRALITY

Elizabeth J. Casey and Christopher D. Wickens
University of Illinois

for

Contracting Officer's Representative
Michael Drillings

BASIC RESEARCH LABORATORY
Michael Kaplan, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

December 1987

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

University of Illinois

Technical review by

Dan Ragland

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC COPY INSPECTED 1

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 87-80	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Visual Display Representation of Multidimensional Systems: The Effect of Information Correlation and Display Integrality		5. TYPE OF REPORT & PERIOD COVERED Interim Report March 85 - March 86
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Elizabeth J. Casey and Christopher D. Wickens		8. CONTRACT OR GRANT NUMBER(s) MDA903-83-K-0255
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology, University of Illinois Champaign, IL 61820		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 20161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE December 1987
		13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) - -		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

- -

18. SUPPLEMENTARY NOTES

Michael Drillings, contracting officer's representative

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dynamic System Information
Cognitive Psychology
Information Displays
Decision Making

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research note provides data on the use of object and schematic face displays to present dynamic multivariate system information. Twelve subjects detected and diagnosed failures in a system whose variables were intercorrelated. Three visual analog displays (a bar graph display, a pentagon, and a schematic face display) represented the system. These displays differed in the degree of integrality of their component features. Detection performance yielded a speed/accuracy tradeoff with little evidence of superiority for any of the displays.

(OVER)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

i SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ARI RESEARCH NOTE 87-80

20. Abstract (continued)

Diagnostic performance showed a superiority for the more separable display, however.

This superiority was attributed to the fact that diagnosis required subjects to focus attention directly on a single attribute, a focussing that benefited from a display which separated the attributes from one another. The results of the study are discussed in a broader context of other studies which looked at the proximity of information. The data also demonstrated the promise of the schematic face display as a means of displaying dynamic system information.

INTRODUCTION

A visual display acts as an interface between a dynamic system and a human operator. Its composition is critical to the performance of the operator in controlling a system and detecting and diagnosing system failures. As the complexity of systems has increased, the amount of information available to the human operator has become overwhelming. Therefore, there is a serious need to optimize the display formats used to present system status information. The operator must be presented with information in a format that requires a minimal amount of mental transformation prior to integrating this information with an already existing internal model of the system. The display format should also allow the operator to respond quickly and accurately when so required.

When acting in a supervisory role, the human formulates a high fidelity internal model of the system. The internal model refers to the human operator's conception of the information structure and serves as a basis for potential actions (Wickens, 1984). A display compatible with the operator's internal model will minimize workload thus allowing faster, more accurate detection and diagnosis. The internal model may vary along several dimensions. Two of these dimensions are the frequency with which the model is updated and the degree to which the representation of the system is spatial and/or verbal (Bainbridge, 1981; Landeweerd, 1979; Wickens and Weingartner, 1985). A third dimension of variability is the perceived degree of <u>integrality</u> of the system variables; in other words, the operator's perception of the relative correlations between the variables or the extent to which critical states of the system are defined by combinations of variables. The present study examines three different methods of graphically representing a dynamic multiattribute system. The hypothesis to be tested is that the integrality of system variables will be best served by more integral displays--objects and faces--than by separated bar graph displays.

One type of graphic representation of multivariate data which has recently received a great deal of attention is the object display which typically represents several variables as attributes of a single geometric object. As an example, consider a polygon formed by connecting the ends of invisible lines which extend out from one point (e.g., Wood, Wise, & Hanes, 1981; Jacob, Egeth, & Bevan, 1976). The length of the imaginary spokes, and therefore the inner angles of the vertices of the polygon represent the values of the system attributes. In addition to giving information about the magnitude of each variable, the overall shape and size of this display can give insight into relationships between the variables. A practical application of the integrated presentation of multivariate data is found in the field of aviation. The contact analog display combines the two variables of roll and pitch into a single, highly schematic representation of the aircraft.

Some of the advantages of the object display over traditional, separate representations of multivariate systems include subjects' familiarity with the objects, the holistic property of object perception by which subjects perceive the overall status of the system, and may process the attributes of a single object in parallel (Kahneman and Treisman, 1984; Kramer, Wickens, Goettl, & Harwood, 1986), and the single frame of reference against which all of the variables can be compared. We hypothesize that the integrated representation provided by the object display will aid the operator in perceiving the relationships among the system variables. This is because a lifetime's experience of dealing with objects and the correlated dimensions of these objects as they are transformed in space, allows the human monitor to associate the integral dimensions that define an object with a correlation between the values along those dimensions. We hypothesize that this association should allow better perception of correlated variables through integral displays. This hypothesis has received some validation in the earlier research of Garner (Garner, 1970; Garner & Fefoldy, 1970). Other research has shown that subjects are particularly sensitive to correlations between variables and thus a display which optimally depicts relational information will be useful to operators of complex, multidimensional systems (Medin, Altom, Edelson & Freko, 1982).

Several empirical studies have been conducted to assess the relative advantages and disadvantages of different displays. In one such study four displays were evaluated: arrays of digits, each digit defining a system variable; glyphs, which portrayed the system variables using the lengths of a series of rays surrounding a circle of fixed size; polygons, the distances from the center to the vertices representing the system variables; and schematic faces, in which each feature delineated a system variable (Jacob, et.al., 1976). Using a card sorting task and a paired associate learning task, Jacob, et. al. demonstrated that people process information from standard displays (such as the arrays of digits) in a "piecemeal, sequential mode which could obscure the recognition of relationships among the individual elements". In contrast, Jacob, et. al., found that the stimuli represented in object displays (the polygon and particularly the face) are processed holistically resulting in easier detection of relationships among variables.

In a series of studies conducted at the Idaho National Engineering Laboratory (INEL) investigators have evaluated the potential use of object displays as Safety Parameter Display Systems (SPDS) in nuclear power plant control rooms (Blackman, Gertman, Gilmore, & Ford, 1983; Danchak, 1981; Gertman, Beckman, Banks, & Petersen, 1982; Petersen, Smith, Banks, & Gertman, 1982). The basic functions of the SPDS include; alerting the operator to the occurrence of abnormal plant conditions, aiding the operator in identifying specific abnormal parameters and assisting the operator in diagnosing plant conditions based on the relative values of parameters. The INEL studies, which have evaluated different object displays in a series of tasks and with several different methodological techniques (psychophysical scaling, multivariate rating scales, checklists and decision analysis), have shown that generally performance with object displays is equivalent or superior to that with more traditional, separate representations of multivariate data. Westinghouse has also proposed and evaluated an object display (polygon) as one of a series of displays to be used in an SPDS (Little & Woods, 1981; Wood, et. al., 1981).

Object displays have also been found useful in presenting a multivariate frame of reference to identify relevant physiological patterns that may delineate the seriousness of medical abnormalities (Siegel, Goldwyn, & Freidman, 1971). Finally, a recent set of investigations carried out at Illinois have suggested conditions that will lead to superiority of the polygon over the bar graph display. Studies by Carswell and Wickens (1984) and Wickens, et. al. (Wickens, Kramer, Barnett, Carswell, Fracker, Goettl, & Harwood, 1985) both indicated that a triangle and rectangle display respectively, offered superior performance when three (or two) pieces of quantitative information needed to be integrated. Another investigation by Kramer, et. al. (1986) found that multivariate graphical information was better integrated when it was presented as a smaller number of more integral objects. More recently Wickens, et. al. (1985) have found that the object display is not universally superior to separated bar graph displays. In fact, when the task required that variables be treated separately from each other, rather than integrated as a single unit, the bar graph display proved superior. Similiarly the bar graph proved to be superior when the task required that attention be focussed on one attribute, to the exclusion of others.

Several investigators have proposed that the holistic perception engendered by schematic faces would be ideal for the presentation of highly related system parameters (Danchak, 1981; Wilkinson, 1981). In one study concerned with the facial representation of multivariate data, the investigator found that the sterotype meaning already present in the faces could be measured and exploited to construct an inherently meaningful display (Jacob, 1978). Thus, in addition to the advantages already cited for object displays, subjects' familiarity with facial expressions appears to provide another dimension which can enhance the perception of multidimensional data. Schematic face displays have been found to be superior to separate numeric presentations of multivariate information in areas as diverse as the financial profile of businesses (Moriarity, 1979), Soviet foreign policy in Sub-Saharan Africa (Wang & Lake, 1978), the evaluation of psychiatric data (Mezzick & Worthington, 1978), and product performance (Hahn, Morgan & Lorensen, 1983).

The program of research we describe here is concerned with explicating the factors that influence the subject's perception, transformation, and response to complex, multi-variate information. This issue is pursued by investigating the conditions under which three different displays (a schematic face, a polygon, and bar graphs) provide an optimal representation of system status information. The following research issues are addressed:

1) Are displays which provide an integrated representation of system parameters (i.e. schematic face and polygon) superior to more traditional displays which present the same information separately (i.e. bar graphs)? Furthermore, does the display format interact with the type of task which the operator is required to perform? Some research has suggested that polygons may be superior to separated meters for detection tasks while meters appear to be optimal for the localization of abnormal variables (Petersen, Banks, & Gertman, 1981; Petersen, et. al., 1982).

2) Does the correlational structure of the system variables interact with the presentation format of the variables? In other words, are different display

formats optimal for systems with different inter-variable correlations? Highly integrated object displays have been proposed to be most useful in situations in which the system parameters are moderately to highly correlated (Wickens, 1984).

3) Do subjects with different degrees of spatial ability adopt different strategies to perform detection and diagnosis tasks? Can we optimize the subjects' performance by presenting system status information in a manner consistent with the subjects' preferred processing strategy?

In this experiment, a temperature process monitoring task was simulated in which subjects were required to detect and then diagnose system failures. Three different displays were evaluated, each representing the correlated variables of the dynamically changing system. The displays, in increasing order of feature integrality were a bar graph display, a polygon, and a schematic face (Wickens, 1984; Jacob, et.al., 1976). The monitored system was presented at two levels of inter-variable correlation, and the spatial/verbal abilities of the subjects were measured.

This task required the integration of the system information: subjects were required to attend to the pattern of correlation between the variables rather than determining whether any of the variables exceeded specified levels of normality. Therefore, it was expected that failure detection and diagnosis performance should increase with display integrality. This effect was predicted to be stronger for the system with the higher correlation between variables.

METHOD

Subjects

Six male and six female University of Illinois graduate and engineering students participated in the study. The subjects, all of whom were right handed, ranged in age from 19 to 25 years. They earned a base rate of $3.50/hour for all four meetings, $.50/day for arriving on time, and bonuses based on performance during the two experimental sessions.

Task

In a simulated setting, the subject monitored a display of the temperatures in five chambers to determine the status of a heating system, as shown schematically in figure 1. The temperature in each chamber was controlled by two sources: a general global furnace which served all five chambers equally, and a local space heater within each chamber which allowed different thermostat settings. The global furnace provided most of the heat (represented as the signal at the top of figure 1 and as S in the temperature equation below each chamber in the same figure) so that the temperatures between the five chambers were correlated over time. However, due to differences in insulation, local temperatures heating needs, etc. (N_i in the figure), the correlations between the chambers were not perfect, as each space heater added its own "noise" to the total temperature variation. A

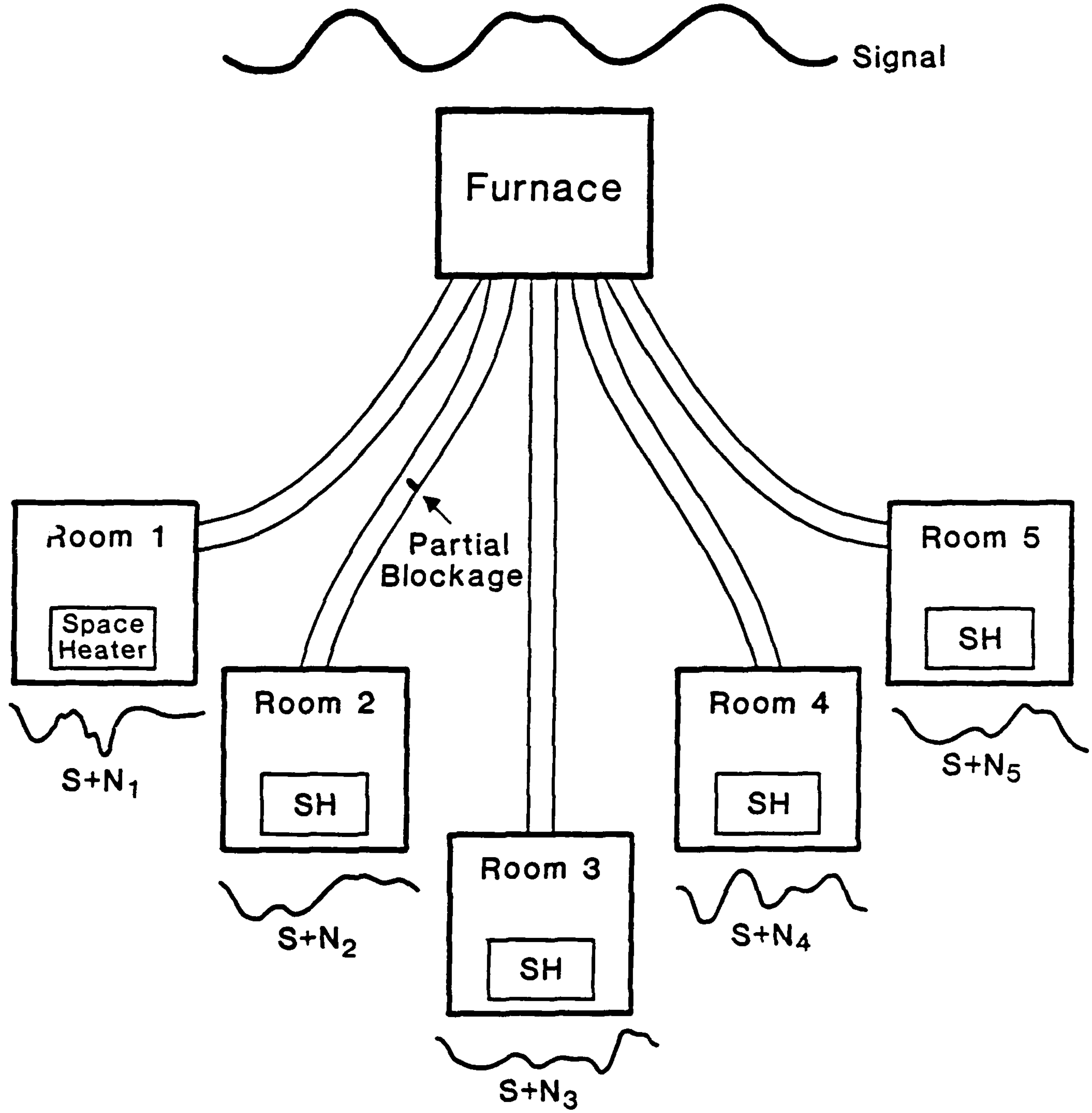


Figure 1. The heating system monitored by subjects. During normal operation the furnace provided the "signal" heat to each of the five chambers. A Space Heater (SH) in each chamber provided additional heating input to the chamber's temperature signal. The curves below the chambers are examples of each chamber's temperature variations over time. During a failure, partially blocked ducting leading to one of the chambers decreased the influence of the furnace input on the chamber's temperature.

failure occurred when a pipe leading from the global furnace to one of the chambers became partially blocked requiring the chamber's space heater to bear a greater burden in heating the chamber. As a result, the temperature in the blocked chamber correlated less with the temperatures of the other chambers than it had before the failure. An example of this change in correlations is demonstrated at the point marked "failure" in figure 2. The temperatures of the other chambers continued to correlate with each other over time.

The subject was instructed to detect failures as quickly and as accurately as possible by pressing a button when the correlation of changes in one signal with changes in the others appeared to drop. Immediately upon correctly detecting the failure, or being told that a failure was missed, the subject diagnosed the location of the failure by pressing a button corresponding to the chamber whose temperature had become less correlated.

Displays

Three visual, analog displays were used to represent the system. In figure 3 temperatures of five chambers are shown using all three displays. The traditional bar graph represented temperatures by the height of the bar with each bar corresponding to one chamber. The object display was a pentagon which represented a chamber's temperature by the distance from the center point to one vertex. Thus, five equal and low temperatures formed a small, equilateral pentagon. The third display was a schematic face. Each of five features represented the temperature in a chamber: ear length, eyebrow angle, eye length, nose

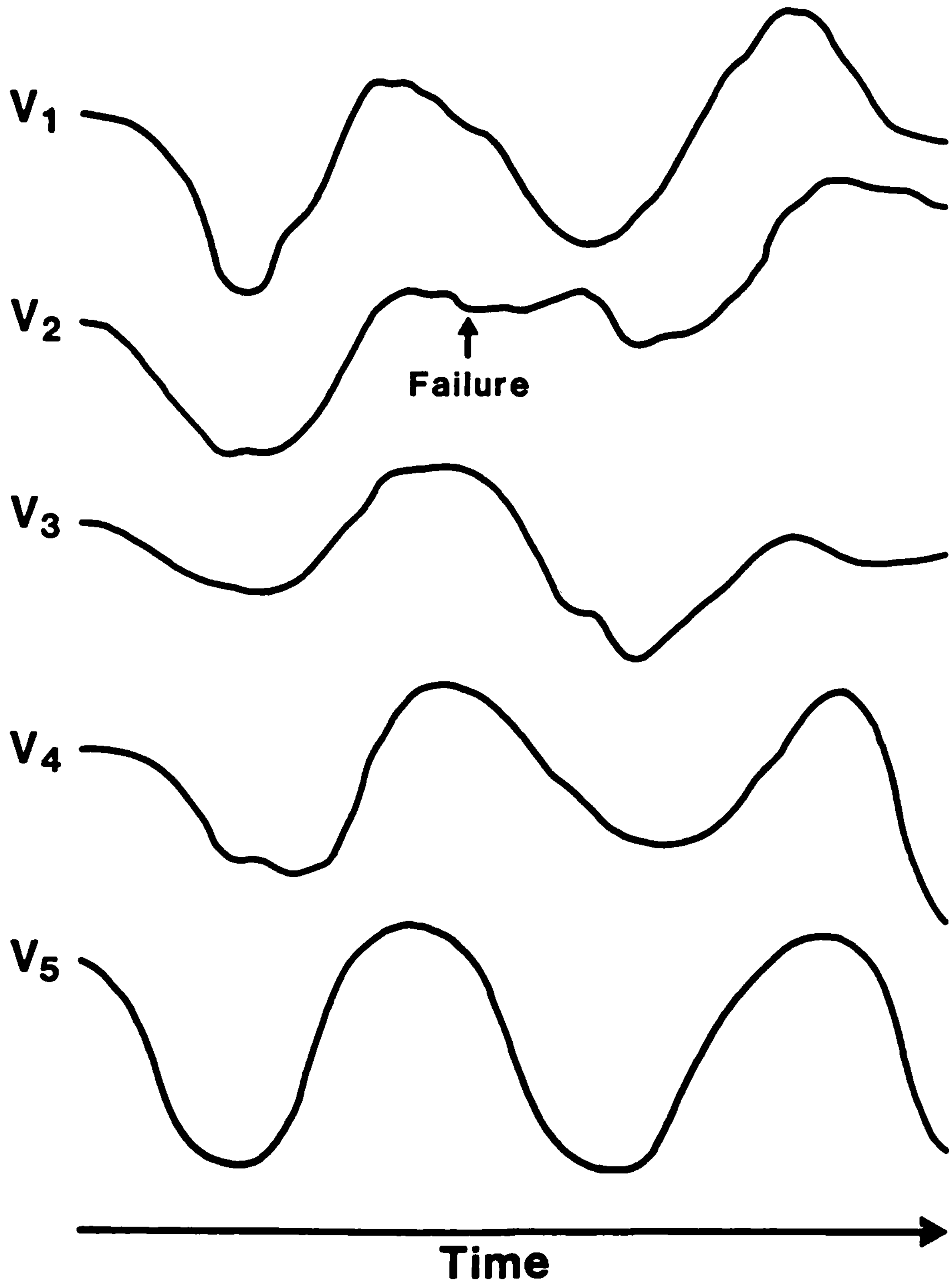


Figure 2. Demonstrated here are representative examples of the temperature of each chamber as they vary over time. The output signals are correlated with each other in the beginning. From the time the failure occurs (in chamber V2 in this case) the correlation of the temperature in the chamber with partially blocked ducting decreases with respect to the temperatures of the other chambers. The correlation of the other four chambers with each other remains constant.

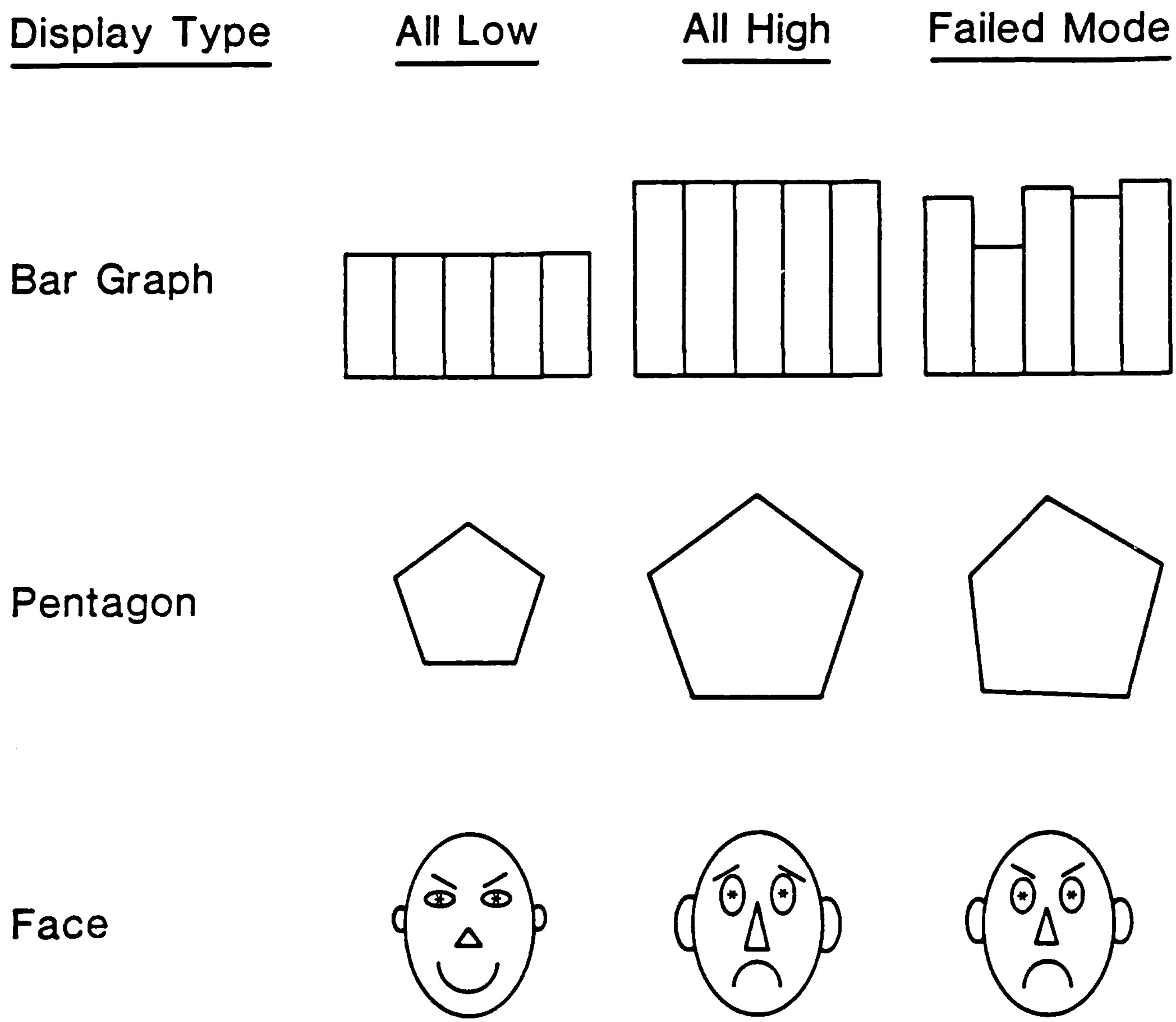


Figure 3. The three displays used to represent the system information. The first two columns of displays are examples of the features with equal temperatures at low and high levels. In the third column the temperature in one of the chambers is lower than the other temperatures. Note that failures were detectable only by discovering changes in the pattern of movement, the change in correlation, between the five display features. Differences in absolute levels of the features at one instant in time were expected under normal operation and did not by themselves indicate system failure.

length, and mouth curvature. As shown in figure 3, a sad face with long features represented all high levels of the variables and the face looked happy (or possibly devious) with short features when the variables were all at low levels. Thus an effort was made to allow the features to correlate over time in the normally operating system, in a manner consistent with the feature correlations over time caused by changes in emotional expression.

The gain on the features within the face and across the three displays was adjusted in order to insure equal discriminability on the basis of results from a preliminary psychophysical scaling study. This study is described in the appendix.

System Dynamics

The displays were updated every half second with temperatures at fifteen discrete levels. The temperature in chamber i was determined using the equation:

$$\text{Temp}(i) = \text{Global} + \text{Deviation}(i) + \text{Error}(i)$$

Global was the contribution of the global furnace. Its value was determined by a random walk, and was the same for each chamber. Sampling from a uniform distribution, the computer determined whether a change from the previous level would occur and, if so, the magnitude of a change using the following probabilities:

no change	30%	
1 level	40%	half increase, half decrease
2 levels	20%	half increase, half decrease
3 levels	10%	half increase, half decrease

Each chamber deviated from the global temperature by a constant value, Deviation(i), for the chamber. This value represented differences in thermostat settings. It ranged from -1 level to +1 level between the five chambers. (Note, in the two "normal" columns of figure 3, the offset deviations are not reflected.)

Error(i) represented the fluctations from moment to moment that were due to differences of insulation, location, etc. For each chamber and each display update the value of Error(i) was chosen randomly from a range of values. Two different baseline levels of correlation were employed; these were established by the magnitude of the error(i), relative to the global signal. The range for the highly correlated system was from -1 to +1 except for the failed chamber whose range was from -3 to +3. For the system with the lower correlation the values ranged from -2 to +2 with fluctuations of the failed chamber ranging from -4 to +4. Operationally, these values produced the mean correlations between system variables of .98 (high) and .93 (low) during normal operations, and correlations of the failed variable with the others of .89 (high) and .78 (low).

When the sum of the three values contributing to a chamber's temperature was less than one or greater than fifteen, the displayed temperature was set to zero or fifteen.

Failures occurred at a random time within an interval of ten and fifty seconds after the beginning of a trial or after diagnosis of the most recent failure. Selection of the chamber whose pipes would be blocked was also random.

Apparatus

The displays were generated by a PDP-11/44 computer on a Hewlett-Packard 1310A display in a small, darkened chamber. The three displays subtended approximately the same degree of visual angle--5° high by 6.5° wide. The lap-held response board featured a stationary joystick with a thumb button for detection response and a five-button box on the right hand side for diagnosis response. The five-button box had four buttons arranged in a semi-circle on top and a fifth button on the front of the box. This arrangement was one that produced an stimulus-response compatible mapping between display features and buttons for both the bar graph and pentagon displays. The mapping was less compatible for the face display. However, a pictorial representation of the appropriate facial feature was presented above each button. The circular pattern was designed based on measurements taken from ten people with a wide range of hand sizes. Data were recorded on magnetic tape.

Procedure

All subjects participated in four sessions: a testing session, a training/practice session, and two experimental testing sessions that were identical with the exception that the order of presentation of the displays and correlation levels were counterbalanced across subjects.

The first meeting was a group sessions which lasted for one hour. The subjects were given four tests of spatial and verbal abilities. The second meeting was a two-hour training session beginning with a detailed verbal, and pictorial description of the system and task. When the subject had a good basic understanding of the experiment, training began in performing the task.

The three types of on-line training were three-minute, modified versions of the actual task. In the first version the word FAILURE appeared on the screen as soon as the system failed followed by an X above the failed chamber. The second type of training allowed the subject to cause a failure by pushing the button corresponding to the chamber whose pipes would be blocked. These two versions gave the subject the opportunity to observe the system in the normal mode to learn the types and amounts of its variation to expect in the system, and then to compare the differences that arose when the system switched to failed mode.

During the third version of training, the subject learned to perform the task without the cues of the first two training types. This was identical to the procedure that would be employed during the experimental session. At each level of training the subject exercised with each display at both correlation levels for a total of 72 minutes of on-line practice.

The third and fourth sessions were the experimental sessions which lasted about two hours each. On each day the subject performed with one correlation level and all three displays with three ten-minute blocks per display. Each block consisted of between eleven and eighteen trials which were structured as shown in figure 4.

```
                                   detection-->|        diagnosis-->|
|____________________|_______|___________|_______|___________|
|<-----10-50 sec---->|<------10 sec----->|<------10sec------>|
Begin                System              Begin                Begin
Trial                Failure             Diagnosis            New
                                                              Trial
```

Figure 4. The time course of events for a single trial. In every trial the system operated normally for at least ten seconds. From that time, a failure could occur at any moment. A failure always occurred within fifty seconds of the beginning of the trial. When a failure began subjects had ten seconds to detect its occurrence. Subjects had another ten seconds from the time of detection (or from the end of the ten second detection interval) to diagnose the failure. A new trial began after the diagnosis was complete or when the ten seconds diagnosis interval was over.

RESULTS

Statistical analyses were conducted on four dependent measures: two for detection and two for diagnosis. Performance on the detection task was measured using average time to detect the failure (measured in milliseconds) and the A-prime measure of sensitivity, a non-parametric measure of signal detection theory which incorporates the probability of a hit and the probability of a false alarm. A-prime was computed for each block of trials using the formula:

$$A' = 1 - 1/4 \; \{ \; [P(FA)/P(H)] + [\; (1-P(H)) \; / \; (1-P(FA)) \;] \; \}$$

(Wickens, 1984) where P(H) = #hits / #trials and P(FA) = #false alarms / #false alarm intervals. A false alarm interval lasted ten seconds since a failure could occur anywhere from ten to fifty seconds into the trial. The number of false alarm intervals in a block was determined to be the total time a signal was not presented divided be ten (the length of a false alarm interval).

Diagnosis performance was measured using average time to diagnose the failure (to the nearest millisecond) and the probability of a correct diagnosis.

All four dependent measures were calculated for each block. These summarized values were analyzed by means of a three-way repeated measures ANOVA. The three within-subject factors were display (bar graphs, pentagon, and face display), correlation level (high vs. low), and block number (3 replications). (Initially subjects were grouped by spatial/verbal abilities, but preliminary analysis revealed that this factor was not a significant source of variation on any of the measures. Hence, spatial/verbal ability was excluded from subsequent analyses, and data for all subjects were pooled.)

The graph at the top of figure 5 illustrates the speed and accuracy failure detection performance measure for each display, cross plotted at each correlation level. The display symbol is positioned at the average levels of the dependent measures collapsed across correlation levels. In this representation, "good" performance (rapid and accurate) are to the upper left, "poor" performance to the lower right, while shifts in a speed versus accuracy set are represented by movement along the positive diagonal. The display manipulation produced a main effect on failure detection latency ($\underline{F}(2,22)=6.3$, $p=.0069$). On the average, subjects detected failures almost half a second faster with the face display ($\underline{M}=3.046$ sec) than with the bar graphs ($\underline{M}=3.496$ sec) or the pentagon ($\underline{M}=3.476$ sec). Accuracy in detection on the other hand was generally highest with the bar graph

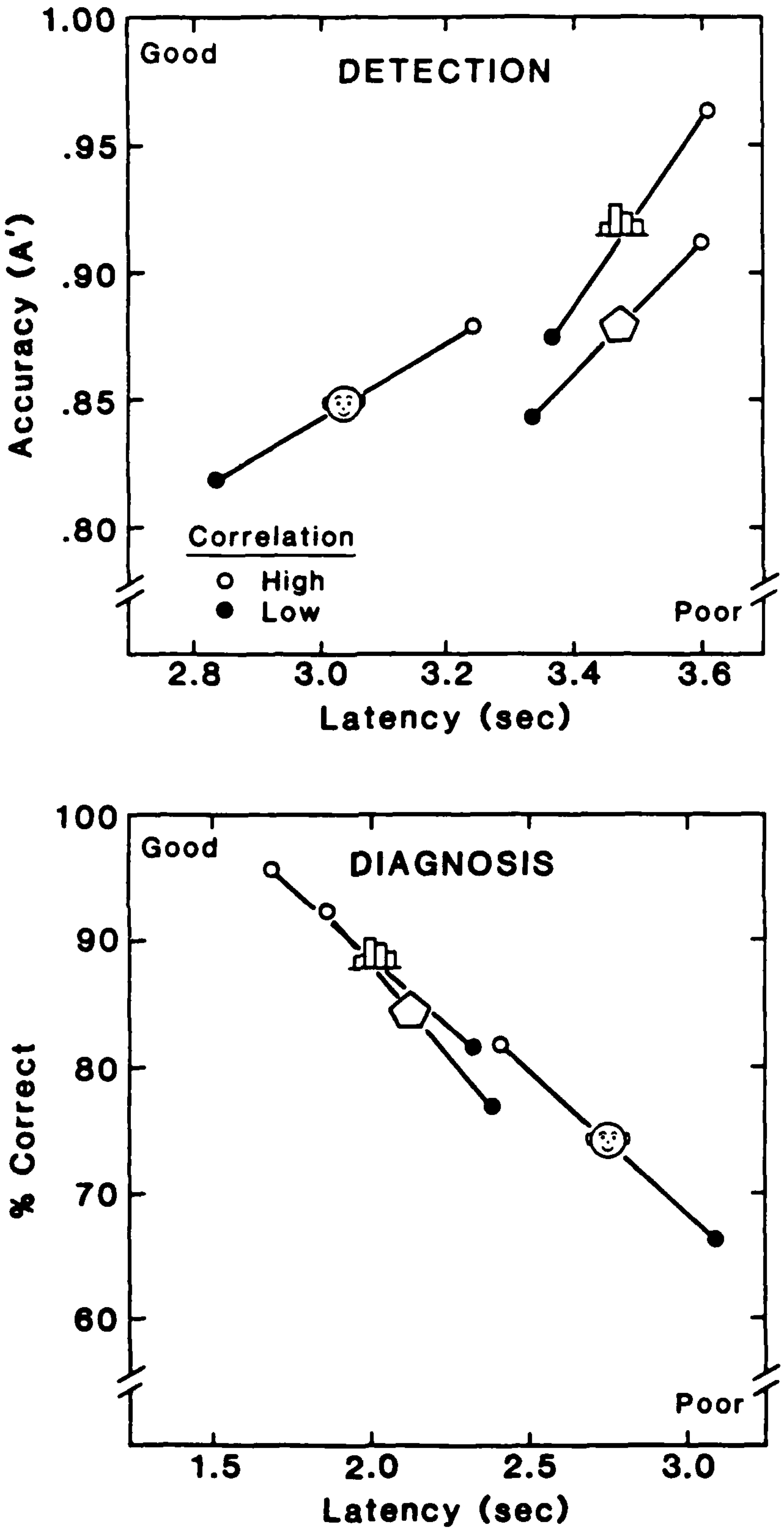


Figure 5. Detection (top graph) and diagnosis (bottom graph) performance results. In both graphs accuracy is plotted against latency for each display at each system correlation level. The endpoints of each line represent high and low correlation levels. The display symbol is positioned at the average of the dependent measures collapsed across correlation levels.

($\underline{M}$=0.919) and lowest with the face display ($\underline{M}$=0.849) ($\underline{F}$(2,22)=20.1,p=.0000).

The Scheffe t-test was used to determine whether performance was better using the bar graph than on the pentagon at the .05 significance level. There were no differences between the two displays in the latency of either detection or diagnosis. However, subjects did perform more accurately on the bar graph when detecting failures (t_s(11)=2.4, p<.05) and when diagnosing the failures (t_s(11)=4.7, p<.05).

It is apparent that the effect of the correlation manipulation was to produce a shift in the speed/accuracy tradeoff, with performance being faster ($\underline{F}$(1,11)=4.7, p=.0528), but less accurate ($\underline{F}$(1,11)=47.4, p=.0000), with a lower correlation between variables.

Both dependent measures of diagnosis performance, shown in the lower graph of figure 5, varied reliably as a function of the display being used. Most of the latency difference was contributed by the face display ($\underline{M}$=2.766 sec) which took over 600 msec longer to diagnose than the bar graphs ($\underline{M}$=2.019 sec) or the pentagon ($\underline{M}$=2.139 sec), ($\underline{F}$(2,22)=9.2, p=.0013). The probability of a correct diagnosis was highest for the bar graph display ($\underline{M}$=0.887), slightly lower for the pentagon ($\underline{M}$=0.846), and another ten percent lower for the face display ($\underline{M}$=0.742), ($\underline{F}$(2,22)=22.7, p=.0000).

Correlation level affected both accuracy and latency: higher correlations produced performance that was both faster ($\underline{F}$(1,11)=7.7, p=.0180) and more accurate ($\underline{F}$(1,11)=45.5, p=.0000).

As shown in figure 6, response time in both the detection task and the diagnosis task did not change significantly as a function of

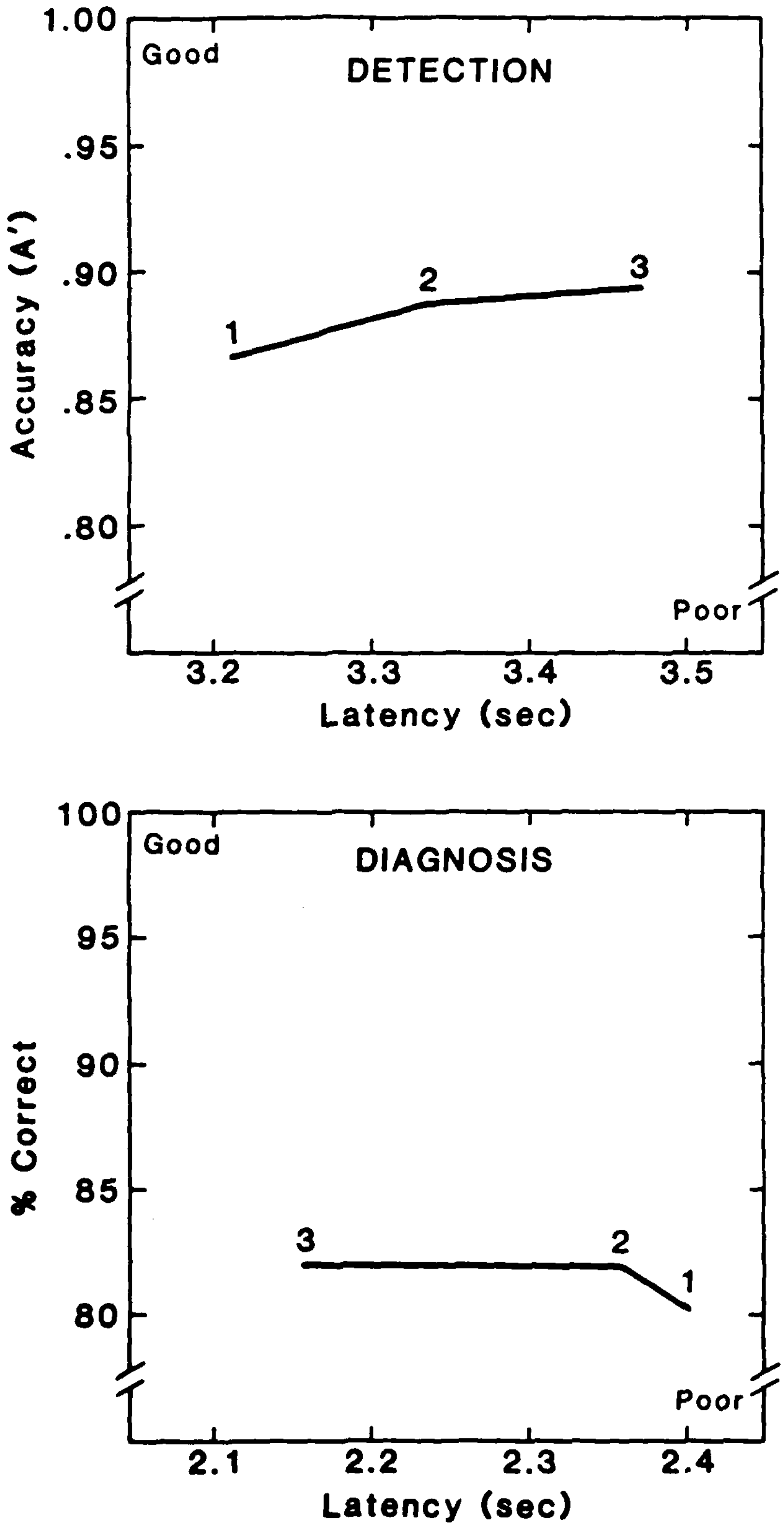


Figure 6. Subject performance as a function of practice.

practice. However, the sensitivity measures for both types of tasks were improved with experience. A-prime, the measure of sensitivity for detection, increased with each block ($\underline{M}(1)=.867$, $\underline{M}(2)=.887$, $\underline{M}(3)=.894$) ($\underline{F}(2,22)=4.9$, $p=.0178$) as did the probability of a correct diagnosis ($\underline{M}(1)=.801$, $\underline{M}(2)=.836$, $\underline{M}(3)=.838$) ($\underline{F}(2,22)=5.7$, $p=.0101$).

There were no significant interactions for any of the measures.

DISCUSSION

The assumption that subjects needed to attend to the patterns of correlations between the variables to accomplish the failure detection and diagnosis tasks led to the hypothesis that performance on both tasks would be superior with the more integrated displays. However, the hypothesis did not prove to be correct.

Detection performance showed a speed/accuracy tradeoff on all three factors: display type, correlation level, and practice. Detections using the schematic face display were faster and less accurate than the other two displays. The least integrated display, the bar graph display, afforded the most accurate but slowest detection responses.

The lower correlation system, which seemed more difficult to the subjects, yielded faster but less accurate failure detection than the more correlated system. The speed/accuracy tradeoff even held for the block factor. With increasing experience, failure detection became more accurate but did not become faster. Generally, with increasing experience, increasing system correlation, and decreasing display integrality, subjects seemed to stress accuracy over speed in their responses in detecting failures.

Diagnosis performance, however, revealed a different pattern of results. In this case for every factor the difference in performance at each level was in overall quality--there were no speed/accuracy tradeoffs. The more integral the features of the display were, the slower and less accurate was performance. The higher correlation level afforded better overall diagnosis performance, and with increasing experience subjects responded more quickly and more accurately.

Because detection and diagnosis show qualitatively different patterns of effects from each other for each of the two manipulated variables, and because the two tasks also manifest different kinds of information processing routines, each will be discussed in turn, before presenting a general theoretical framework for interpreting the effects of display integrality.

As noted, the data for the detection task indicate a speed/accuracy tradeoff across both the display manipulation and the manipulation of correlation. Some conditions (face display, low correlation) yielded fast but inaccurate responses, while others (bar graph, high correlation) yielded responses that were slow and accurate. Two interpretations however may be offered to this pattern of data. On the one hand, it is possible that there is really little difference in the efficiency or effectiveness of detection performance across these conditions. They differed only in the "set" for speed versus accuracy with which subjects chose to operate. On the other hand, there may have been some fundamental limitations of the less accurate conditions (low correlation and face display) that prohibited subjects from attaining a higher level of accuracy, even by prolonging latency. Such a phenomenon has been observed elsewhere in signal detection experiments that have shown negative speed/accuracy tradeoffs (i.e. less accurate performance with longer latencies; Vickers, 1970; Welford, 1976). Hence, in these "difficult" conditions subjects may have decided that since there is no advantage to accuracy to be gained by waiting longer, a rapid response might as well be given.

The critical test necessary to choose between these two hypotheses would have been to induce subjects to adopt different speed/accuracy sets <u>within</u> a condition. Thus, according to the first hypothesis, a request for subjects to adopt a conservative criterion setting (slower and more accurate), in the face (or low correlation) condition would have "moved" performance to a level at which it had the same speed and accuracy as the more integrated displays (and high correlation conditions). According to the second hypothesis, the request would have prolonged latency, with no increase in accuracy. Unfortunately, since this manipulation was never performed, the hypothesis cannot be tested. There is, at most, some converging evidence for the second hypothesis, based upon the subjective reports that the face display and low correlation conditions were somewhat more difficult.

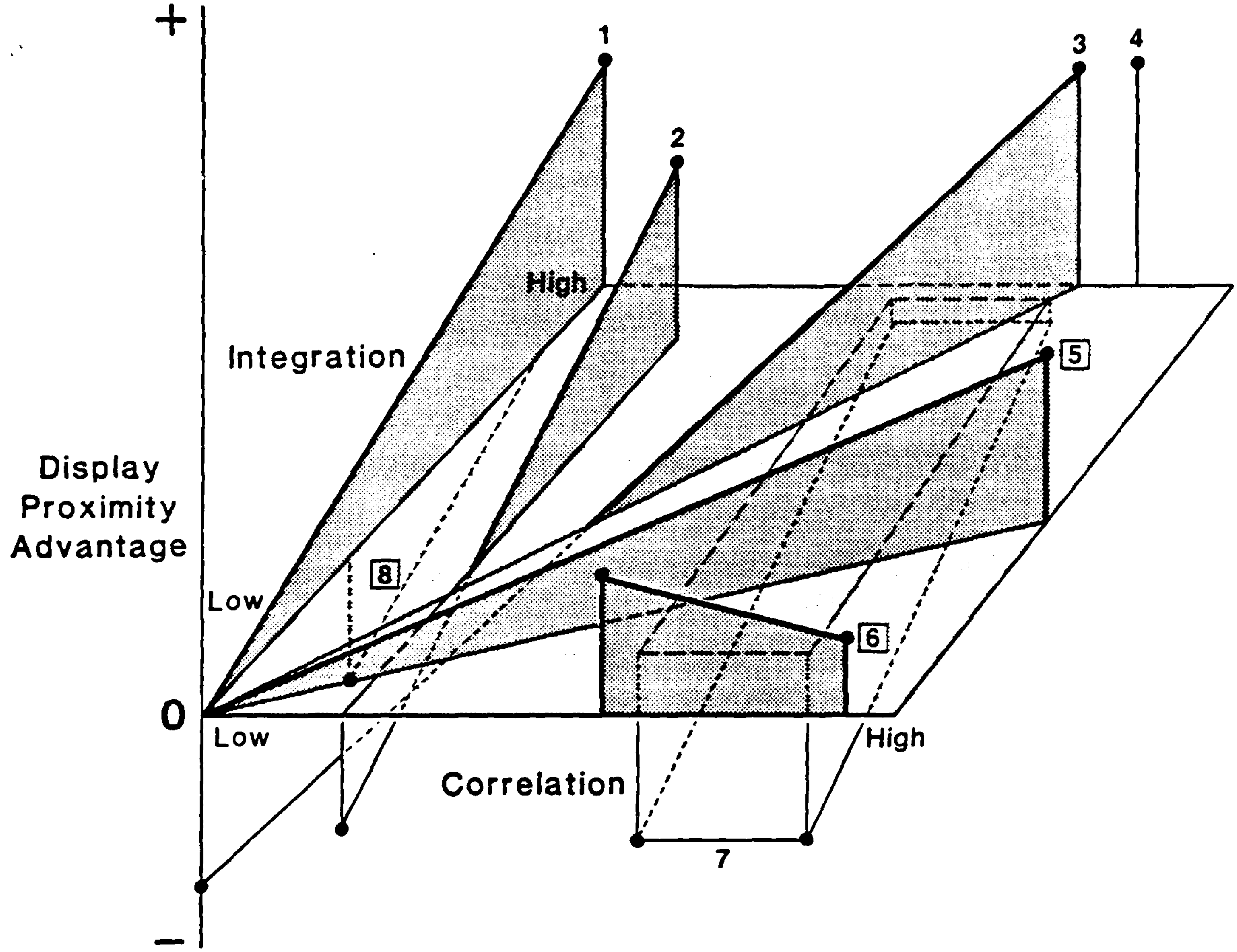
Unlike detection, there was no ambiguity regarding the relative ordering of merit of diagnosis across conditions. The lower correlation condition and the face display were reliably worse along both performance dimensions. The polygon display was reliably less accurate than the least integral bar graph display.

An overall summary of the results of both tasks would state that more integral displays (the polygon and particularly the face), lead to poorer diagnosis performance, but have a less harmful effect (or perhaps no effect) on detection performance. This result, which for diagnosis was not originally predicted, must be rectified with other results in the literature that have obtained an advantage for the object or face display in diagnosis sorts of tasks (e.g., Jacob, et. al., 1976; Carswell & Wickens, 1984; Wickens, et. al., 1985). The difference between the present result, and those that have found superiority of integral displays appears to lie in a difference in the nature of the diagnosis required, and in the contrast between focussed attention and information integration. In both the paradigms used by Jacob, et. al. and by Carswell and Wickens, there was a complex mapping of variable values to diagnostic state. That is, each diagnostic state required the simultaneous consideration of more than a single variable. This condition of <u>information integration</u> is one that we have argued elsewhere, is best served by integral displays (e.g., Carswell & Wickens, 1984; Wickens, et. al., 1985; Wickens, 1986). In contrast, the diagnosis task used in the present experiment imposed a 1-to-1 mapping between variable and diagnostic state, therefore imposing more of a requirement to <u>focus</u> attention on a single variable; that is, treating the failed variable as separate and unique from its neighbors. It would be plausible to argue here that such conditions are not favored by integral displays. In fact this view is consistent with another condition reported by Wickens, et. al., (1985) in their comparison of bar graphs with object displays. When their task was modified to require a separate 1-to-1 mapping for each variable to a different response, the integrality advantage was actually reversed, and performance became superior with the bar graph display. The results of Peterson, et. al.'s (1982) investigation comparing separated meters with integrated stars also are somewhat consistent. When fault location was required (a 1-to-1 mapping), performance was best with the separated meters, and poorer with the integrated star, although they found that the separated bar graph display did not perform at the level of the meters.

Note that the interpretation offered above is consistent with the results of the detection task in the current data as well as those of Peterson, et. al. (1982). Detection, like the diagnosis tasks of Carswell and Wickens, requires a many-to-1 mapping, in which a larger number of displayed variables must be mapped into a smaller number of cognitive and response states. This condition is by definition one of information integration which should benefit from the more integral displays. In the investigations of Peterson, et. al. and Carswell and Wickens, this benefit was present. In the current study there was no absolute benefit to the integral face and object displays in detection -- only a reduction of their cost, relative to the diagnosis condition.

The absence of an absolute benefit for object displays in the present detection data could probably be attributed to some other inherent disadvantage to the face and polyon display in the current study. Subjects, for example, voiced some complaint that the large visual angle subtended by the moving parts of the polygon display was fatiguing. Also the selection of features for the face display may not have been optimal. On the one hand, two of the features chosen, ear length and nose length, are not features that naturally change dynamically within a face. A second possible source of incompatibility with the face display relates to the concept of heterogeneity. The features of the face, while holistically integrated, are also heterogeneous, each one having a different physical appearance, meaning, and emotional content. This heterogeneity, of course, is in contrast to the points on the pentagon, or the five bar graphs which are homogeneous. Yet, the particular system which provided the context for the scenario was also a homogeneous one, with all five variables having the same semantic meaning. Thus, a "compatibility of homogeneity" between display and system variables that was present for the pentagon and bar graph display, was absent for the face display. Both of these issues, the role of constant versus changing features in the face, and the issues of homogeneity, are presently under investigation in follow-on studies in our laboratory.

The broader context of display integrality and information processing, within which the current results may be interpreted is presented in figure 7. On the ordinate of the figure is represented what is termed a "Display Proximity Advantage" or D.P.A. That is, an advantage in a particular experiment for displays that are "close" or integral, such as the face or the object, over displays that are separated. Negative values of this D.P.A. are those such as observed for the diagnosis task in the present experiment. The two dimensions of the abscissa in this three-dimensional representation reflect the two task/cognitive variables that were hypothesized to influence the D.P.A.: information integration and information correlation. To the extent that a D.P.A. is modulated by either of these cognitive variables, the concept of display integrality moves beyond the perception domain, and makes it necessary to invoke central processing concepts such as the internal model or mental integration. This is



1. Barnett (Wickens et. al., 1985, exp. 5).
2. Banks et. al., 1982.
3. Carswell & Wickens (Wickens et. al., 1985).
4. Jacob et.al., 1976.
5. Garner, 1971.
6. Kramer et. al., 1985.
7. Casey & Wickens, present exp.
8. Goettl (Wickens et. al., 1985, exp.2).

Figure 7. This figure portrays the Display Proximity Advantage (D.P.A.) as a function of the amount of correlation between the displayed values, and the degree to which those values must be integrated. Conditions for which integration is low are those that require either focussed attention on one source of information, or independent processing of several sources (i.e., dual or multi-task processing). Each experiment is designated by a number, identified in the legend above. Solid lines and planes indicate a Display Proximity Advantage, and thus lie above the plane of the surface. Dashed lines and open planes indicate experiments or conditions with a disadvantage to proximate displays. They thus depict negative values below the origin of this three-dimensional representation.

because a display principle that is purely perceptual in its characteristics, should be unaffected by the later cognitive processes required of the displayed information. The relevance of the two abscissa dimensions to cognitive, rather than perceptual phenomena are as follows: task integration requires that the joint consequence of all stimuli must be taken into account before a response is made -- that is, some sort of mental operation must be carried out on the stimulus as an ensemble, rather than either focussing attention only on a single stimuli, or dividing attention between stimuli but processing each independently of the others. Stimulus correlation need not have cognitive implications. However, to the extent that correlated stimuli are processed better than orthogonal stimuli (a valid assumption in decision-making research; Moray, 1981; Ebbeson & Konecki, 1981), then some higher level of cognitive processing must be operating to extract the presence of this correlation, and thereby use it to advantage.

Figure 7 then shows the effects of task integration and correlation on the D.P.A. Each vertical line on the graph indicates a pair of conditions in which display proximity has been manipulated, by one form or another. A vertical "slice" is an experiment in which display proximity has been manipulated orthogonally with another variable that effects either the degree of correlation, or the amount of integration required. A vertical solid has manipulated both proximity and correlation orthogonally. For example, since the current experiment varied the degree of integration between detection (higher) and diagnosis (lower) using information sources which were always correlated (but whose correlation varied), this experiment is represented by a solid whose position on the plane is as labelled.

Some experiments have contrasted the D.P.A. in two conditions that have varied, in a confounded manner, both in the correlation between their inputs and in the amount of information integration required. Hence, the "planes" defined by their results are oriented at an angle to the two axes.

While it is difficult to draw any conclusion with absolute certainty from the data shown in this representation, two general trends appear to be noteworthy. (1) There is a general tendency for the D.P.A. to increase (or a display proximity disadvantage to dissipate), as tasks require more information integration, or less divided and focussed attention. That is, the contours "slope" upward from the front of the figure to the back. (2) The effect of correlation on the D.P.A. appears to be substantially less. There is little trend in D.P.A. from the left of the figure to the right as correlation between displayed variables increases.

In summary, it is hoped that this representation will provide the foundation for a theory-based means of predicting the circumstances in which more "integrated" displays may or may not be employed to advantage over more separated formats. Such guidelines will not,of course, be absolute. For example, as we have noted in the present data, there may be numerous influences on the merits of a face display that exist independent of the degree of integration required (e.g., the assignement of features to variables or the heterogeneity of variables). However, such a framework does, it is hoped, establish the foundation for a theory of display integration.

APPENDIX

Scaling Study

It is important in any comparison of visual displays to determine whether the superiority of a given visual display can be accounted for by perceptual factors. Therefore, in order to properly compare these displays we first had to ensure that it was equally difficult to perceive a change in a system variable regardless of the display or display feature on which that variable was represented. Thus, a psychophysical scaling study was conducted on five displays. (Originally meters and glyphs were being studied in addition to the other three displays. These were omitted in the correlation/display study because it proved impossible to perform the task in that study with those displays.)

Ten college students participated in the experiment. All were right-handed with normal or corrected-to-normal vision. During the two hour session the subjects performed the task with each of the five displays. The subject's task was to decide whether two sequentially presented displays matched or mismatched. The importance of both speed and accuracy was emphasized. Subjects pressed one response button if the displays matched and another if they did not. In a single block of trials only one of the five variables on a display was to be attended by the subject. The other four variables remained at constant levels. Each of the five features of the face varied in different blocks.

For each display and feature, ten equidistant levels were defined. The "standard" (S1) was always either at level 5 or at level 11. The comparison stimulus (S2) varied as follows:

STANDARD	PERCENT TRIALS	COMPARISON	PERCENT TRIALS
		UP TWO LEVELS	12.5
LEVEL 5	50.0	UP ONE LEVEL	12.5
OR		SAME	50.0
LEVEL 11	50.0	DOWN ONE LEVEL	12.5
		DOWN TWO LEVELS	12.5

In total there were nine blocks, five for the face and one each for the other four displays. Before each block of 160 trials, the subjects had fifteen practice trials. Experimental blocks and response buttons were counterbalanced across subjects.

The amount of time required to decide whether two displays matched or mismatched was affected by the type of display being judged ($\underline{F}(8,72)=4.3$, $p<.01$). The order of displays from fastest to slowest was the meters ($\underline{M}=318$ msec), bar graphs ($\underline{M}=334$ msec), polygon ($\underline{M}=342$ msec), glyphs ($\underline{M}=373$ msec) and schematic face ($\underline{M}=394$ msec). The amount of time required to compare different facial features ranged from 375 msec for the eyebrows to 417 msec for the mouth. RT was also influenced by stimulus level. Slight mismatches took longer to respond to than matches and more obvious mismatches ($\underline{F}(4,36)=18.3$, $p<.01$). There was an interaction between display type and stimulus level such that RT performance with meters and bar graphs was not differentially affected by stimulus level ($\underline{F}(32, 288)=1.7$, $p<.01$). Error rate generally followed the same pattern as RT with larger error rates being associated with longer RTs.

The results of this scaling study were used to adjust the magnitudes of the physical changes of the display components. The ranges of variations were scaled to be psychophysically equivalent. Thus, any differences in performance among displays are not attributable to perceptual factors.

REFERENCES

Bainbridge, L. (1981). Mathematical equations or processing routines? In Human Detection and Diagnosis of System Failures (J. Rasmussen and W.B. Rouse, eds.), New York: Plenum Press.

Blackman, H.S., Gertman, D.I., Gilmore, W.E. and Ford, R.E. (1983). Noninteractive simulator evaluation for CRT-generated displays. Idaho National Engineering Laboratory, NUREG/CR-3556.

Carswell, C.M. and Wickens C.D. (1984). Stimulus integrality in displays of system input-output relationships: A failure detection study. In Proceedings. 28th Annual Meeting of the Human Factors Society (E. Alluisi ed.), Santa Monica, CA: Human Factors Press.

Danchak, M.M. (1981). Techniques for displaying data on cathode ray tubes with applications to nuclear process control. Idaho National Engineering Laboratory, NUREG/CR-1944.

Ebbeson, E.D. and Konecki, V. (1980). On external validity in decision-making research. In Cognitive processes in choice and decision making (T. Wallsten, ed.), Hillside, NJ: Erlbaum Asssociates.

Garner, W.R. (1970). The stimulus in information processing. American Psychologist, 25, 350-358.

Garner, W.R., & Fefoldy, G.L. (1970). Integrality of stimulus dimension in various types of information processing. Cognitive Psychology, 1, 225-241.

Gertman, D.I., Beckman, H.S., Banks, W.W. and Petersen, R.J. (1982). CRT display evaluation: The multidimensional rating of CRT generated displays. Idaho National Engineering Laboratory, NUREG/CR-2942.

Hahn, G.J., Morgan, C.B. and Lorensen, A.E. (1983). Color face plots for displaying product performance. IEEE Computer Graphics and Applications, 3, 23-29.

Jacob, J.K. (1978). Facial representation of multivariate data. In Graphical Representation of Multivariate Data (P.C.C. Wang, ed.), New York: Academic Press.

Jacob, J.K., Egeth, H.E., and Bevan, W. (1976). The face as a data display. Human Factors, 18, 189-200.

Kahneman, D., and Treisman, A. (1984). Changing views of attention and automaticity. In Varieties of Attention (R. Parasuraman and R. Davies, eds.), New York: Academic Press.

Kramer, A., Wickens, C., Goettl, B., and Harwood, K. (1986). Extraction and comprehension of data relations: an application of the display proximity principle. In Proceedings of the Symposium on Psychophysiology in the Department of Defense (R. Ginnett, ed.), United States Air Force Academy.

Landeweerd, J.A. (1979). Internal representation of a process, fault diagnosis and correction. Ergonomics, 22, 1343-1351.

Little, J.L. and Woods, D.D. (1981). A design methodology for the man-machine interface for nuclear power plant emergency response facilities. Proceedings of the Joint American Nuclear Society and Taiwan Nuclear Society Conference on Human Factors and Safety. Taipei, Taiwan.

Medin, D.L., Altom, M.W., Edelson, S.M., and Freko, D. (1982). Correlated Symptoms and Simulated Medical Classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 37-50.

Mezzick, J.E. and Worthington, D.R.L. (1978). A comparison of graphical representations of multivariate psychiatric diagnostic data. In Graphical Representation of Multivariate Data (P.C.C. Wang, ed.), New York: Academic Press.

Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of errors in man-machine systems. In Human detection and diagnosis of system failures (J. Rasmussen and W. Rouse, eds.), New York: Plenum Press.

Moriatrity, S. (1979). Communicating finanical information through multidimensional graphics. Journal of Accounting Research, 17, 205-224.

Petersen, R.J., Banks, W.W. and Gertman, D.I. (1981). Performance based evaluation of graphic displays for nuclear power plant control rooms. Proceedings of the Conference on Human Factors in Computer Systems, Gaithersburg, Maryland.

Petersen, R.J., Smith, R.L., Banks, W.W. and Gertman, D.I. (1982). An empirical examination of evaluation methods for computer generated displays: Psychophysics. Idaho National Engineering Laboratory, NUREG/CR-2942.

Siegel, J.H., Goldwyn, R.M. and Freidman, H.P. (1971). Pattern and process in the evolution of septic shock. Surgery, 70, 232-245.

Vickers,D. (1970). Evidence for an accumulator model of psychophysical discrimination. Ergonomics, 13, 37-58.

Wang, P.C.C. and Lake, G.E. (1978). Application of graphical multivariate techniques in policy sciences. In Graphical Representation of Multivariate Data (P.C.C. Wang, ed.), New York: Academic Press.

Welford, A.T., (1976). Skilled performance: Perceptual and motor skills. Glenview, IL:Scott, Foresman.

Wickens, C.D. (1984). Engineering Psychology and Human Performance. Columbus, Ohio: Charles E. Merrill Publishing Company.

Wickens, C.D. (1986). Attention. In Human Factors in Psychology (P. Hancock, ed.), Amsterdam: North Holland.

Wickens, C.D., Kramer, A., Barnett, B., Carswell, M., Fracker, L., Goettl, B., Harwood, K. (1985). Display/cognitive interface: the effect of information integration requirements on display formatting for C3 displays. Tech. Report EPL-85-3/AFHRL-RADC-85-1, Engineering-Psychology Research Lab, University of Illinois, Urbana.

Wickens, C.D. and Weingartner, A. (1985). Process control monitoring: the effects of spatial and verbal ability and concurrent task demand. In Proceedings of Second Mid-Central Ergonomic/Human Factors Conference--Trends in Ergonimics/Human Factors (R. Eberts and C.G. Eberts, eds.), Vol. II., Purdue University.

Wilkinson, L. (1981). An experimental evaluation of multivariate graphical point representations. Proceedings of the Conference on Human Factors in Computer Systems, Gaithersburg, Maryland.

Wood, D., Wise, J. and Hanes, L. (1981). An evaluation of nuclear power plant safety parameter display systems. In Proceedings of the 25th Annual Meeting of the Human Factors Society (R.C. Sugarman, ed.), Santa Monica, California: Human Factors Press.